REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
09-05-2016	Final Report	29-May-2015 - 28-Feb-2016

4. TITLE AND SUBTITLE
Final Report: Variations on Bayesian Prediction and Inference

5a. CONTRACT NUMBER
W911NF-15-1-0154

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Ryan Martin

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Illinois - Chicago
MB 502, M/C 551
809 S. Marshfield Avenue
Chicago, IL 60612 -4305

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
66552-MA-II.4

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
A Bayesian approach, based on updating prior information in light of new observations, via Bayes's formula, has both nice intuition and strong theoretical support. However, in some applications, there are some roadblocks to carrying out the Bayesian analysis as usual. This ARO-sponsored project considered two general types of these settings. First, for the problem of predicting a future observation, a flexible Bayesian model is available but is too computationally expensive to implement when data are streaming and fast prediction is required. Second, for the problem of inference about an unknown parameter, the Bayesian approach requires a full probability

15. SUBJECT TERMS
Bayes; prediction; likelihood-free; Gibbs posterior

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Ryan Martin
UU	UU	UU	UU		19b. TELEPHONE NUMBER 312-413-2156

Report Title

Final Report: Variations on Bayesian Prediction and Inference

ABSTRACT

A Bayesian approach, based on updating prior information in light of new observations, via Bayes's formula, has both nice intuition and strong theoretical support. However, in some applications, there are some roadblocks to carrying out the Bayesian analysis as usual. This ARO-sponsored project considered two general types of these settings. First, for the problem of predicting a future observation, a flexible Bayesian model is available but is too computationally expensive to implement when data are streaming and fast prediction is required. Second, for the problem of inference about an unknown parameter, the Bayesian approach requires a full probability model/likelihood which can be an obstacle; indeed, there are applications where no likelihood function is available or, even in cases where a likelihood is or could be made available, it may be preferable not to use it because it depends on nuisance parameters, etc. Work done in this project addresses both the prediction and inference problems, expanding the range of applications of the Bayesian approach to help meet the Army's needs.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

My student, Nick Syring, has presented his work done in the paper "Scaling the Gibbs posterior credible regions" in our department's Statistics Seminar. He will also be presenting his work in the paper "Image boundary detection via a Gibbs posterior" at the end of May 2016 at the Spring Research Conference being held at Illinois Institute of Technology. My collaborator (Stephen Walker) has presented, on several occasions, our work in the paper "On recursive Bayesian predictive distributions".

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received		Paper
05/06/2016	1.00	Nick Syring, Ryan Martin. Likelihood-free Bayesian inference on the minimum clinically important difference, arXiv:1501.01840 (09 2015)
05/06/2016	2.00	Nick Syring, Ryan Martin. Scaling the Gibbs posterior credible regions, ArXiv: 1509.00922 (09 2015)
05/06/2016	3.00	P. Richard Hahn, Ryan Martin, Stephen G. Walker. On recursive Bayesian predictive distributions, arXiv:1508.07448 (12 2015)
TOTAL:	**3**	

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Nick Syring	1.00	
FTE Equivalent:	**1.00**	
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):.....	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Technology Transfer

ARO-STIR — Final Project Report
Contract No. W911NF-15-1-0154
"Variations on Bayesian prediction and inference"

Ryan Martin
Department of Mathematics, Statistics, and Computer Science
University of Illinois at Chicago
rgmartin@uic.edu
www.math.uic.edu/~rgmartin

May 9, 2016

Abstract

The familiar Bayesian framework, where observed data is used to update prior information, via Bayes's formula, has many desirable features. This project aims to address shortcomings of this Bayesian approach in two essential problems, namely, prediction and inference. First, for the prediction problem, the Monte Carlo computation required to obtain a genuine Bayesian predictive distribution can be too slow for use with streaming data, and a new recursive estimator of the Bayesian predictive distribution is proposed which is both fast to compute and has desirable theoretical properties. Second, for the inference problem, there are cases where a full probability model for all the unknowns is not available and/or is not desirable, so there is a need for "likelihood-free" Bayesian inference. New tools are developed to address various theoretical and computational questions related to the use of so-called Gibbs models for such problems.

1 Statement of the problems

1.1 Problem 1: prediction

Consider the problem where data, $Y_1, \ldots, Y_n, \ldots$ are iid from some distribution with a density function p. After observing $Y_1, \ldots, Y_n$, the goal is to predict the next observation Y_{n+1}. More specifically, we desire a rule that converts observations $(Y_1, \ldots, Y_n)$ into a density function p_n which is, in some sense, our "best guess" of the distribution of Y_{n+1}. This object is usually called a *predictive density*. A standard way to construct a predictive distribution is by taking a Bayesian approach. That is, start with a prior distribution Π for the unknown density p supported on a subset $\mathbb{P}$ of all density functions. Then the Bayesian approach constructs a posterior distribution for p, given $(Y_1, \ldots, Y_n)$, denoted

by Π_n, that satisfies

$$\Pi_n(B) = \frac{\int_B \prod_{i=1}^n p(Y_i)\,\Pi(dp)}{\int_{\mathbb{P}} \prod_{i=1}^n p(Y_i)\,\Pi(dp)}, \quad B \subseteq \mathbb{P}.$$

This allows for a natural construction of a predictive distribution. Given the posterior Π_n, define the predictive density p_n for Y_{n+1}, given $(Y_1, \ldots, Y_n)$, as the posterior mean density, i.e.,

$$p_n(y) = \int p(y)\,\Pi_n(dp).$$

This Bayesian predictive distribution is logically quite reasonable and, except for some rather unusual cases, will have good theoretical properties. The downside, however, is that, for standard choices of prior Π, such as a Dirichlet process mixture model, the Monte Carlo computations required to evaluate p_n are somewhat costly. More concerning is that updating p_n to p_{n+1} when Y_{n+1} is observed requires that one redo all the Monte Carlo computations, so online Bayesian prediction is not possible. The first goal of this project is to develop a recursive approximation to the Bayesian predictive update, one that does not require Monte Carlo computations, and has desirable theoretical properties.

1.2 Problem 2: inference

Consider the same general setup as in the previous section, but now the main interest is in some feature $\theta = \theta(p)$ of the distribution p. Here θ could be a finite-dimensional vector, e.g., a set of moments, or could be a function, like in nonparametric regression. From a non-Bayesian point of view, depending on the setup, it may be possible to produce an estimator or other kinds of inference for θ without specifying a model for p. For example, one can easily construct an estimator and an asymptotically correct confidence interval for a quantile without specifying a model for the underlying distribution p. A Bayesian approach, on the other hand, requires a likelihood function which, in turn, requires a model for p. There is no conceptual difficulty to specify a prior Π for p, get a posterior Π_n based on data $(Y_1, \ldots, Y_n)$, and then get a corresponding posterior distribution for θ via marginalization. The question is why spend the time and resources to specify a prior distribution for the full distribution p and carry out the potentially non-trivial Monte Carlo computations to evaluate the posterior distribution of p, when the intention is to marginalize over everything but θ? A second goal of this project is to develop theoretical and computational tools that will allow users to carry out a Bayesian analysis working only on the interest parameter θ, avoiding the unwanted tasks of prior specification and Monte Carlo computations over the nuisance parameter space.

2 Summary of main results

2.1 Problem 1: prediction

2.1.1 Background

An important special case of the general setup in Section 1.1 is that of a Dirichlet process mixture model for the density p; see, for example, Lo (1984) and Escobar and West

(1995). That is, the prior Π models the density p as a mixture

$$\int K(y \mid u)\, G(du),$$

where $K(y \mid u)$ is a kernel, and the mixing distribution G is modeled by a Dirichlet process distribution (Ferguson 1973). By now there is a substantial literature on Monte Carlo methods to fit this model and compute the corresponding predictive; see MacEachern and Müller (1998), Neal (2000), Walker (2007), and Kalli et al (2011). However, these methods do not allow for a mapping from the previous predictive and a new observation, (p_{n-1}, Y_n), to the updated predictive, p_n. The burden of requiring Monte Carlo methods motivated Newton and Zhang (1999) and Newton (2002) to consider an approximation of the full Dirichlet process mixture model, called *predictive recursion*, which produced a recursive estimator of the mixing measure G. Extensive study of Newton's recursive estimator is carried out in Tokdar et al (2009) and Martin and Tokdar (2009, 2011). This estimator is fast and easy to compute but, despite its name, it does not directly target the predictive density—numerical integration is needed to evaluate normalizing constants, etc. The goal here is to develop a version of Newton's method that directly targets the predictive and, therefore, does not require numerical integration. Interestingly, to accomplish this goal, we will need to make unexpected use of copulas.

2.1.2 Results

As a candidate for the predictive update, one with a multiplicative structure, consider a bivariate function $k(y, y')$ that satisfies

$$p_n(y) = p_{n-1}(y) k(y, Y_n).$$

Therefore,

$$k(y, Y_n) = \frac{p_n(y)}{p_{n-1}(y)}$$

which is symmetric in (y, Y_n), since

$$k(y, Y_n) = \frac{\int p(y)\, p(Y_n)\, \Pi_{n-1}(dp)}{\int p(y)\, \Pi_{n-1}(dp) \int p(Y_n)\, \Pi_{n-1}(dp)}. \tag{1}$$

Hahn et al (2015) show that the function $k(y, Y_n)$ in (1) is a bivariate copula density function (Nelsen 1999); that is, for some symmetric copula density c_n, which depends on the sample only through the sample size, we have

$$k(y, Y_n) = c_n\big(P_{n-1}(y), P_{n-1}(Y_n)\big) \tag{2}$$

where $c_n(u, v) = c_n(v, u)$ is a symmetric copula density, and P_{n-1} is the distribution function corresponding to the predictive density p_{n-1}.

We can now write the update $(p_{n-1}, y_n) \mapsto p_n$ as

$$p_n(y) = c_n(P_{n-1}(y), P_{n-1}(y_n))\, p_{n-1}(y) \tag{3}$$

and for each Bayesian model there is a unique sequence c_n. Now (3) allows for the direct update of the predictive and moreover it can be seen that all one needs to direct a

sequence of predictive densities is to define a sequence of copula functions c_n, the key to which is that $c_n \to 1$ as $n \to \infty$, i.e., the sequence of copula converges to the independent copula as the sample size increases.

Unfortunately, it is only possible to write down the corresponding copula density c_n for relatively simple parametric models. We can, however, follow Newton's approach and build a recursive approximation based on what is known about the Dirichlet process mixture model. In particular, fix an initial guess P_0, with density p_0, a sequence of weights $(\alpha_n) \subset (0,1)$, and a correlation parameter $\rho \in (0,1)$. Then, sequentially compute

$$p_n(y) = (1-\alpha_n)\, p_{n-1}(y) + \alpha_n\, p_{n-1}(y)\, h_\rho(P_{n-1}(y), P_{n-1}(Y_n)), \quad n \geq 1. \tag{4}$$

where h_ρ is the Gaussian copula density

$$h_\rho(u,v) = \frac{\mathsf{N}_2(\Phi^{-1}(u), \Phi^{-1}(v) \mid 0, 1, \rho)}{\mathsf{N}(\Phi^{-1}(u) \mid 0,1)\mathsf{N}(\Phi^{-1}(v) \mid 0,1)}.$$

Under some mild regularity conditions on the true density, $p^\star$, if the weight sequence (α_n) satisfies $\sum_n \alpha_n = \infty$ and $\sum_n \alpha_n^2 < \infty$, then the recursive predictive p_n in (4) is a consistent estimator, i.e., $p_n \to p^\star$ almost surely, as $n \to \infty$, with respect to the L_1 norm. In addition to the theoretical consistency results, numerical simulations reveal that the recursive estimator is fast and easy to compute and also highly accurate.

2.1.3 Further developments

The paper (Hahn et al 2015) that contains these results is currently under revision for the *Journal of the American Statistical Association* and, though we cannot be sure of the outcome at this point, we expect that it will ultimately be accepted there. As part of our revision, we intend to consider some possible extensions of the results summarized above, including some dependent-data cases, as well as some multivariate prediction problems. In part through the work on recursive density estimation for this project, Stephen Walker and I developed some new ideas for solving classical inverse problems, such as solving Fredholm equations, using statistical ideas/methods. We recently learned that this new project will be supported, in part, by the National Science Foundation.

2.2 Problem 2: inference

2.2.1 Background

There are a number of statistical inference problems that are not generally formulated via a full probability model. Perhaps the most important example of these is when the goal is inference on quantiles, especially, quantile regression. The usual non-Bayesian approach to quantile regression, as discussed in Koenker (2005), formulates the problem as one of optimization, i.e., minimize a measure of the empirical risk, and there is a rich theory of M-estimation on which to build this framework. A Bayesian approach, on the other hand, is less straightforward because one is used to building a posterior distribution based on a prior and a likelihood, but here there is no likelihood. Bayesian quantile regression has received some attention in the literature, e.g., Yu and Moyeed (2001) and Sriram et al (2013), and these also make use of the same empirical risk, but the formulation is in

terms of a misspecified likelihood. This perspective does not make clear, however, how to approach other problems when no likelihood is present.

A concrete motivating example for this work comes from a medical application in Hedayat et al (2015). One objective of a clinical trial is assessing the efficacy of a treatment, but statistical significance alone does not necessarily imply efficacy. For instance, a study with high power may detect statistically significant differences that do not translate to practical differences noticeable by the patients. As a result, a cutoff value different than a statistical critical value is desired that would separate patients with clinically significant responses from those patients without a clinically significant response. This cutoff is deemed the "minimum clinically important difference" (MCID).

Let $Y \in \{-1, 1\}$ denote the patient reported outcome with 1 meaning that the treatment was effective and -1 meaning that the treatment was not effective. Let X be a continuous diagnostic measure taken on each patient. The MCID, denoted by $\theta^\star$, satisfies

$$\mathsf{P}\{Y \neq \operatorname{sign}(X - \theta)\} = \min_\theta \mathsf{P}\{Y \neq \operatorname{sign}(X - \theta)\},$$

i.e., θ minimizes, over θ, the probability that $\operatorname{sign}(X - \theta^\star)$ disagrees with Y. Here, P is with respect to the joint distribution of (X, Y). Since

$$\mathsf{P}\{Y \neq \operatorname{sign}(X - \theta)\} = \mathsf{E}\Big\{\frac{1 - Y \operatorname{sign}(X - \theta)}{2}\Big\},$$

and the latter can be interpreted as an expected loss, the MCID is understood as an expected loss minimizer, as discussed above. Note that the MCID is defined without a model for the distribution of (X, Y), but estimation can proceed without it; in fact, Hedayat et al (2015) proceed to estimate the MCID by minimizing an empirical version of the above expression:

$$R_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{1 - Y_i \operatorname{sign}(X_i - \theta)}{2}. \tag{5}$$

How could one solve this problem from a Bayesian perspective? One could introduce a probability model, and priors for the model parameters, and compute the corresponding posterior, but the results surely would depend on the particular choice of model, which may not be correct. Moreover, the probability model likely would depend on other nuisance parameters, not just on θ alone, and this complicates the problem even further. It would be desirable to produce a posterior distribution for θ directly, without worrying about if the model is wrong or dealing with nuisance parameters.

Bissiri et al (2016) recently showed that a logical Bayesian prior-to-posterior update can be carried out in cases where only a loss function, and not a likelihood, is available. Their update corresponds to the so-called Gibbs posterior which has been used occasionally in the statistics and machine learning literature. In particular, suppose the inference problem is defined by a loss function that admits a corresponding empirical risk function $R_n(\theta)$ like that in (5). Then the Gibbs posterior distribution satisfies

$$\Pi_n(d\theta) \propto e^{-wnR_n(\theta)}\, \Pi(d\theta), \tag{6}$$

where Π is a prior distribution for θ and $w > 0$ is a scale parameter. Note that the Gibbs posterior in (6) involves no nuisance parameters, so investments in prior specification

and posterior computations can be focused on the parameter of interest. Compare this to standard Bayesian methods that require a prior and posterior for everything. Although the Gibbs posterior does depend on w, it often maintains good theoretical properties over a range of w values, and its finite-sample performance may even be improved by using data-dependent choices of w; see Syring and Martin (2015ab).

2.2.2 Results

There are three main issues addressed in this project. First, to see the benefit of the Gibbs approach, important examples need to be identified. Second, with motivation coming from some interesting and important applications, what are the general properties that Gibbs posterior distribution would satisfy, in particular, asymptotic concentration properties? Third, there will be computational challenges to be overcome in order to implement the proposed methodology. In what follows, we will summarize the results obtained on each of these points.

In terms of applications, the two previously mentioned, namely, MCID and quantile regression, are important ones, and these are addressed specifically in Syring and Martin (2015ab). The MCID problem is essentially a classification problem, so there are a variety of extensions of this that would be of interest. In particular, Hedayat et al (2015) investigate an interesting *personalized MCID* problem where the MCID is actually a function of some patient-specific characteristics; some preliminary work on a Bayesian approach to this problem has been completed, but more details remain to be filled in. There are a few other interesting applications that we have started working on, and we discuss these in more detail in Section 2.2.3 below.

In terms of general theory, we have shown in Syring and Martin (2015b) that, under very general conditions, the Gibbs posterior distribution will concentrate around the true $\theta^\star$ at the same rate as the corresponding M-estimator, i.e., the estimator obtained by minimizing the empirical risk function $R_n(\theta)$. In addition, we have shown that any reasonable choice of the scale w, even one that depends suitably on n, will not affect the Gibbs posterior concentration rate. Extensions of these concentration rate results to handle adaptivity are currently being completed; see Section 2.2.3 below.

In terms of computation, there are several issues. In simple finite-dimensional problems, such as MCID or quantile regression, it is relatively straightforward to compute the Gibbs posterior using the standard Metropolis–Hastings machinery. On the other hand, in nonparametric problems with infinite-dimensional parameters, things are not so easy. Indeed, the standard Bayesian computational methods employed in these problems often take advantage of some structure—conjugacy or otherwise—to simplify things but, of course, the empirical risk function is unlikely to have any such structures. We have been addressing this issue case-by-case so far, but progress has been made.

Finally, one last point that fits in with all three of the categories mentioned above is the choice of the scaling parameter w in (6). It does not have an effect on the Gibbs posterior concentration rate, but it does have an effect on the finite sample performance. What criterion should be used to make a good choice of w? In Syring and Martin (2015b), we proposed to choose w such that the corresponding Gibbs posterior credible regions are calibrated, i.e., so that the credible regions are also confidence regions. The algorithm to make this choice uses a combination of bootstrap and stochastic approximation. Sur-

prisingly, this choice of w leads to exact nominal coverage probability in our simulation examples, for each fixed sample size, not just asymptotically.

2.2.3 Further developments

The papers listed above (Syring and Martin 2015ab) are both currently under revision, and we expect to get these back into the review pipeline shortly. The results are good but, so far, it seems we have not been able to make clear what are the main contributions. In addition to these two papers, we have one more that is near completion. This latter paper considers the problem of nonparametric estimation of the boundary of an image based on noisy intensity measurements at pixels. This is a particularly challenging problem for the classical Bayesians because they have to specify the intensity distribution inside and outside the image. This introduces nuisance parameters and potential bias if the models are misspecified. He have developed a Gibbs posterior approach that is easy to compute—based on a suitable reversible jump Markov chain Monte Carlo method—and enjoys adaptive optimal concentration rates. We are particularly excited about this work, and we expect it will be completed by the end of May 2016.

There are many other potential applications to be explored within this Gibbs model framework. One that has caught my attention is in the Lévy process models that often appear in financial applications, i.e., models for continuous-time/streaming data that allow for the possibility that the sample paths have jump discontinuities. For such models, often the quantity of interest is the Lévy density, which is what controls the size and frequency of jumps. Even if the model is fully specified, neither writing down/evaluating the likelihood function nor simulating the data-generating process is straightforward, so Bayesian methods have not been used. A Gibbs model, however, is fairly straightforward to implement and can directly target the Lévy density, avoiding nuisance parameters, etc. This is an exciting problem and I hope to make some progress this summer.

3 Student training

This STIR award provided support for one of my students, Nick Syring, during Summer and Fall of 2015. The work completed while Nick was supported by the ARO will make up a significant portion of his doctoral thesis. I have taken a new job at North Carolina State University, starting in Fall 2016, and Nick will be joining me there. His thesis will surely acknowledge the gracious support from the ARO.

4 References

Bissiri, P., Holmes, C., and Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society Series B*, to apperar.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

Hahn, P. R., Martin, R., and Walker, S. G. (2015). On recursive predictive distributions. Unpublished manuscript, arXiv:1508.07448.

Hedayat, S., Wang, J., and Xu, T. (2015). Minimum clinically important difference in medical studies. *Biometrics* **71**, 33–41.

Kalli, M., Griffin, J. E., Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing* **21**, 93–105.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates I. Density estimates. *Annals of Statistics* **12**, 351–357.

MacEachern, S. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 223–238.

Martin, R. and Tokdar, S. T. (2009). Asymptotic properties of predictive recursion: robustness and rate of convergence. *Electronic Journal of Statistics* **3**, 1455–1472.

Martin, R. and Tokdar, S. T. (2011). Semiparametric inference in mixture models with predictive recursion marginal likelihood. *Biometrika* **98**, 567–582.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265.

Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer.

Newton, M. A. and Zhang, Y. (1999). A recursive algorithm for nonparametric analysis with missing data. *Biometrika* **86**, 15–26.

Newton, M. A. (2002). On a nonparametric recursive estimator of the mixing distribution. *Sankhya* **64**, 306–322.

Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2013). Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian Analysis* **8**, 479–504.

Syring, N. and Martin, R. (2015a). Likelihood-free Bayesian inference on the minimum clinically important difference. Unpublished manuscript, arXiv:1501.01840.

Syring, N. and Martin, R. (2015b). Scaling the Gibbs posterior credible regions. Unpublished manuscript, arXiv:1509.00922.

Tokdar, S. T., Martin, R., and Ghosh, J. K. (2009). Consistency of a recursive estimate of mixing distributions. *Annals of Statistics* **37**, 2502–2522.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics–Simulation and Computation* **36**, 45–54.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters* **54**, 437–447.